EEU44C04 / CS4031 / CS7NS3 / EEP55C27
Next Generation Networks

Other M/M/… queues

Nicola Marchetti
nicola.marchetti@tcd.ie

Other Markov queuing systems

- A number of other queuing systems are similar to the M/M/1 queue
  - Poisson arrival process
  - Exponential service times
- Similar systems:
  - M/M/1/N: finite buffer (system capacity = N)
  - M/M/m: m servers (rather than 1)
  - M/M/∞: infinite number of servers (no queuing)
  - M/M/m/m: m servers without queuing

M/M/1/N queue

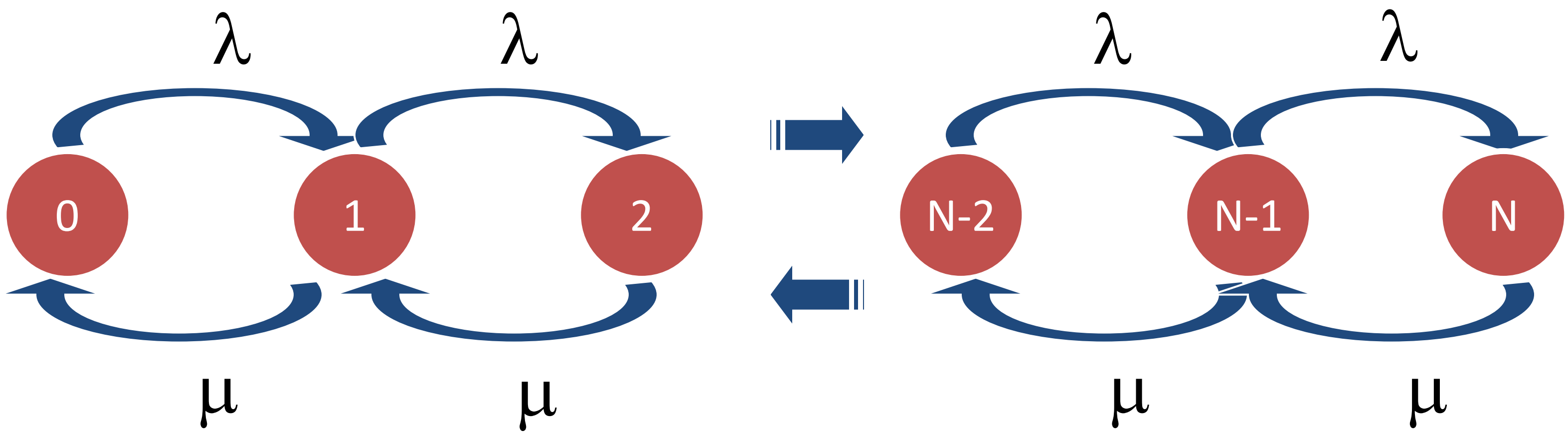
- Poisson arrival process
- Exponential service times
- Single server
- Finite capacity of *N* customers in the system

finite capacity *N*

M/M/1/N Markov chain model

- Terminates with N in the system
- Local balance equations yield …

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0, \quad 0 \le n \le N$$



M/M/1/N steady-state solution

- For $\rho = \lambda/\mu < 1$

$$P_n = \frac{(1-\rho)\rho^n}{1-\rho^{N+1}}, 0 \le n \le N$$

$$\bar{n} = E[n] = \sum_{n=0}^{N} nP_n = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}$$

Blocking probability

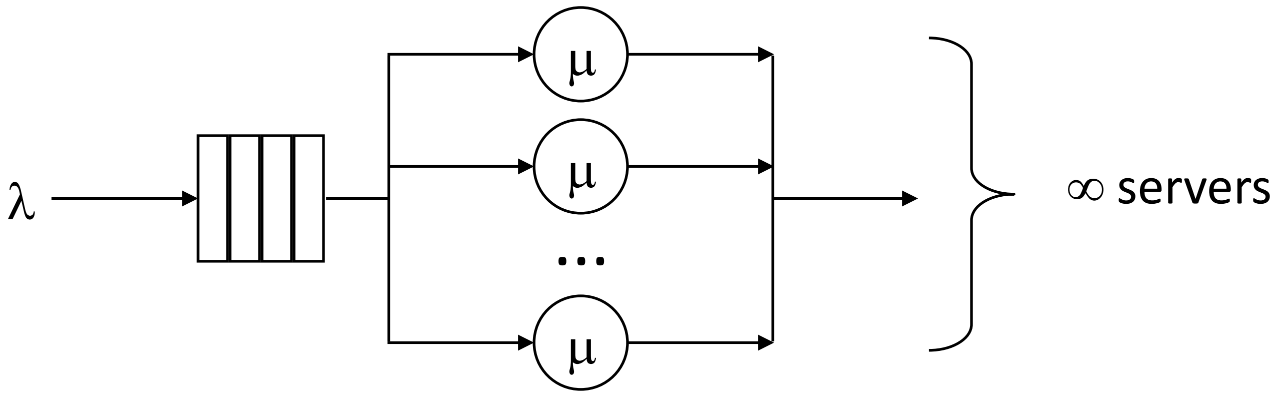
- Blocking probability is an important performance measure for systems with finite capacity

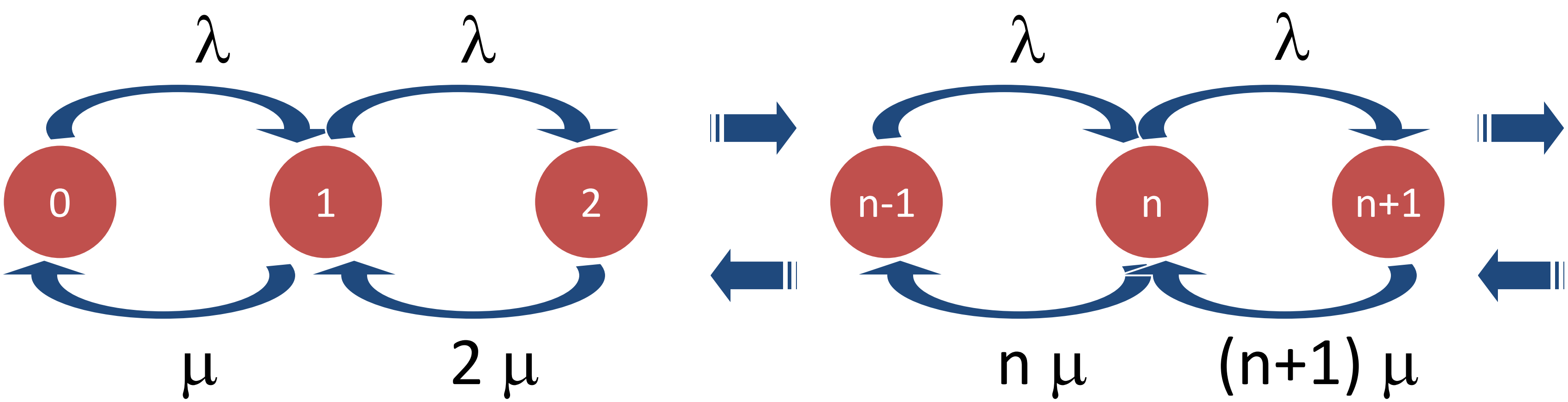
- The **blocking probability** is the probability that an arriving customer will find the system full, i.e., that there are N customers in the system
    - Blocking probability: P_N
    - Rate of rejected customers: λP_N

M/M/∞

- The M/M/∞ queue is a special case of the M/M/m queue with an infinite number of servers (m=∞)
  - Every arriving customer is put into service
    - E.g., a swimming pool, or other self-service activity
  - **No queuing delay**, only service time

M/M/∞ Markov chain model





M/M/∞ results (1)

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \frac{P_0}{n!} = \left(\frac{\lambda}{\mu}\right)^n \frac{e^{-\lambda/\mu}}{n!}, \quad n = 0,1,\ldots$$

M/M/∞ results (2)

- Valid for $0 \leq \lambda/\mu < \infty$ (i.e., always stable!)
  - There is an infinite number of servers

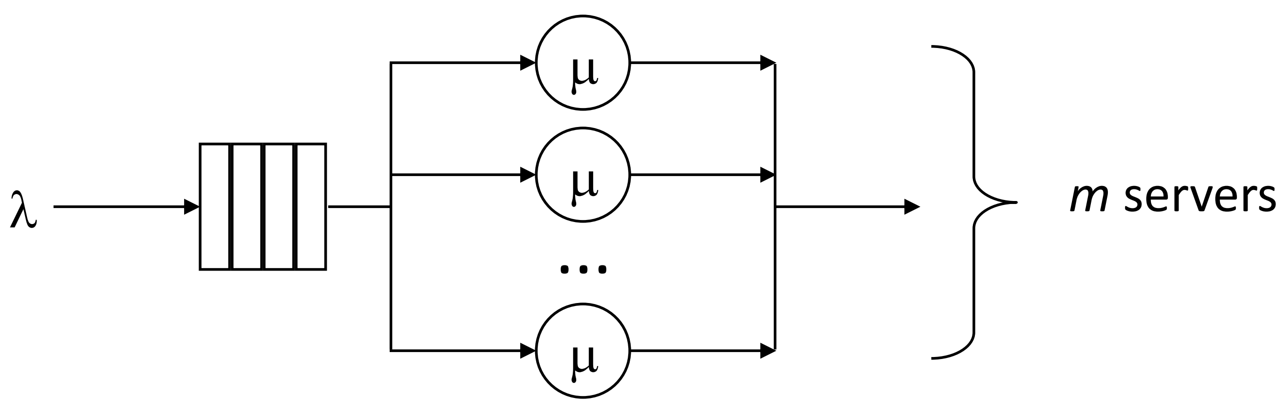
- The number in the *system* for the M/M/∞ is Poisson distributed with parameter λ/μ

- The expected number in the system is $N = \lambda/\mu$

- Little's Theorem gives the average delay per customer
  - $T = N/\lambda = 1/\mu$
  - This is as expected. It is the service time… there is no queuing

M/M/m queue

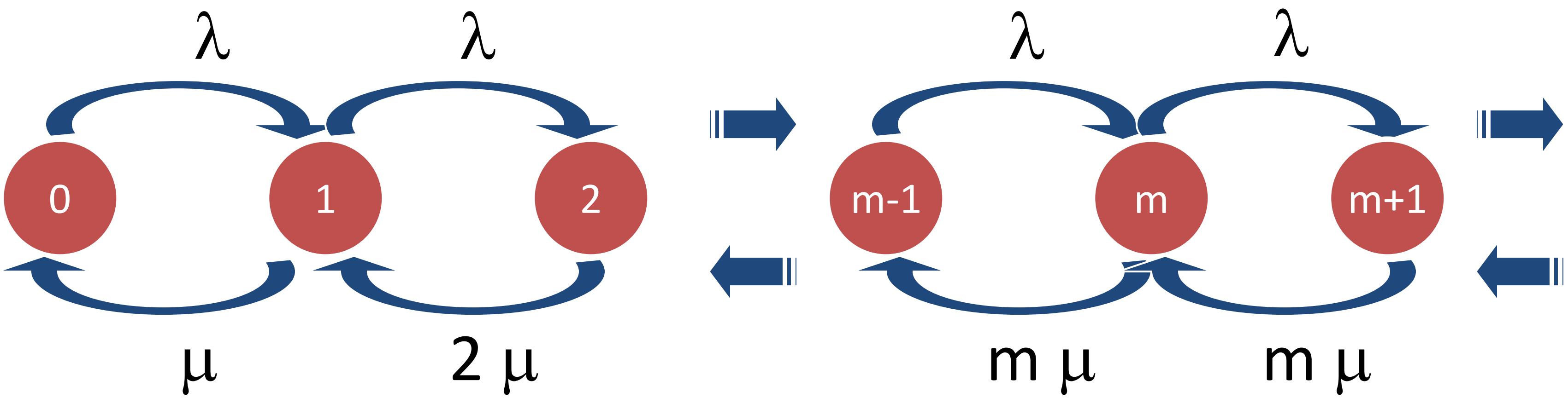
- Identical to the M/M/1 queue except that there are m servers
  - (E.g., m channels)
- Any arriving customer (packet) is routed to any server that is available; if none available, the packet enters a single queue



M/M/m Markov chain model

$$\lambda(n) = \lambda$$

$$\mu(n) = \begin{cases} n\mu & 0 \le n \le m \\ m\mu & n \ge m \end{cases}$$



M/M/m probability of queuing

- "Erlang C" Formula (widely used in telephony and circuit switching systems)

$$P_Q = \frac{(m\rho)^m}{m!(1-\rho)} P_0 = \left(\frac{1}{m!}\right)\left(\frac{\lambda}{\mu}\right)^m \frac{1}{(1-\rho)} P_0$$

- We define P_0 later, in the problems section

- **Note**: we define ρ as $\lambda/(m\mu)$ for an M/M/m queue

M/M/m expected number in the queue

- The expected number of customers in queue, but not in service is …

$$N_Q = P_Q \frac{\rho}{(1-\rho)}$$

- **Note**: we define ρ as $\lambda/(m\mu)$ for an M/M/m queue

M/M/m other results (1)

- Little's Law gives the expected time in queue …

$$W = \frac{N_Q}{\lambda} = \frac{\rho P_Q}{\lambda(1-\rho)} = \frac{P_Q}{m\mu - \lambda}$$

- Expected delay per customer (queuing and service) …

$$T = \frac{1}{\mu} + W = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda}$$

service delay
(= 1/service rate)

queuing delay

M/M/m other results (2)

- Expected number in the system (using Little's Law) …

$$N = \lambda T = \frac{\lambda}{\mu} + \frac{\lambda P_Q}{m\mu - \lambda} = m\rho + \frac{\rho P_Q}{1-\rho}$$

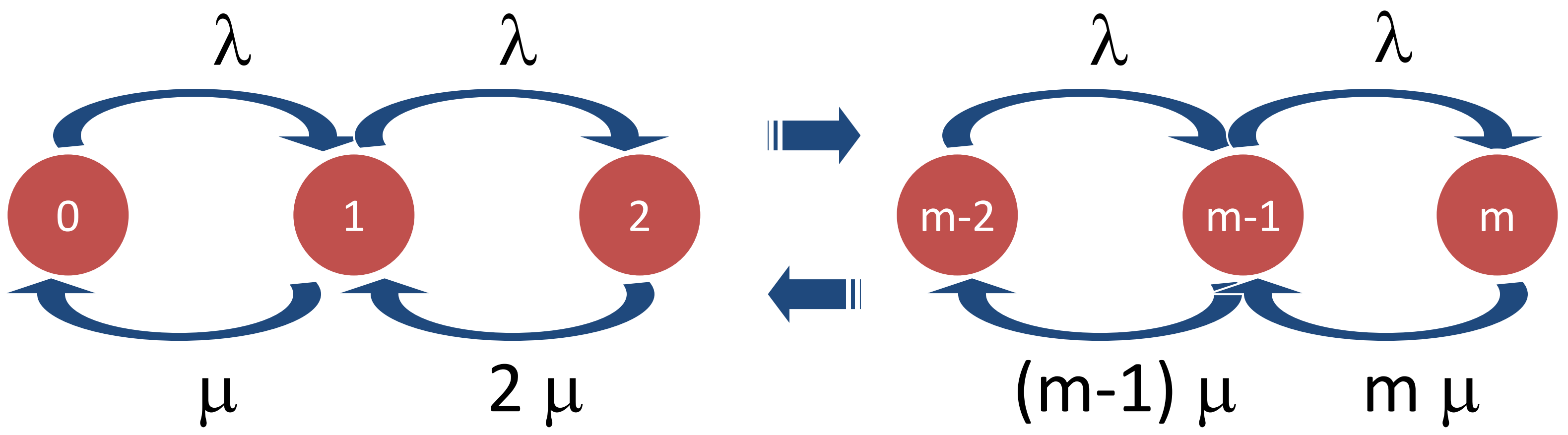
M/M/m/m queue

- The M/M/m/m queue is known as the "m-server loss system"
  - There are *m* servers (like the M/M/m queue)
  - The last "m" in M/M/m/m denotes that the system capacity is limited to *m* customers
  - Queuing is not allowed, so a customer that arrives at a system with all servers busy is **lost**

M/M/m/m Markov chain model

$$\lambda(n) = \lambda \quad n = 0,1,\ldots,m-1$$

$$\mu(n) = n\mu \quad n = 1,2,\ldots,m$$



M/M/m/m blocking probability (1)

- The blocking probability is the probability that an arriving packet will find all servers (circuits) busy and will, therefore, be lost or "blocked"
- The blocking probability is P_m, the probability that all m servers are busy

$$P_m = \left(\frac{\lambda}{\mu}\right)^m \frac{P_0}{m!} = \frac{\left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!}}{\sum_{n=0}^{m}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}} = B$$

"B" = Blocking

M/M/m/m blocking probability (2)

- Can be shown to hold for the M/G/m/m system (general service times)

- Also known as the Erlang B formula or the Erlang-Loss formula
  - Widely used to evaluate blocking probability in telephone systems

Problem

Suppose that messages arrive according to a Poisson process at a rate of one message every 4 ms, and that message transmission times are exponentially distributed with mean 3 ms.
The system maintains buffers for 4 messages, including the one being served.

What is the blocking probability?

What is the average number of messages in the system?

- **M/M/1/N**, with N=4
- **arrival rate** $\lambda = 1/4$ msg/msec
- **service rate** $\mu = 1/3$ msg/msec
- $\rho = \lambda / \mu = 3/4$

What is the blocking probability?

$$P_N = P_4 = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}} = \frac{\left(1-\dfrac{3}{4}\right)\left(\dfrac{3}{4}\right)^4}{1-\left(\dfrac{3}{4}\right)^5} \approx 0.1$$

What is the average number of messages in the system?

$$E[n] = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} = \frac{3/4}{1-3/4} - \frac{(4+1)(3/4)^5}{1-(3/4)^5} \approx 1.44$$

See slide 5

Problem

In a self-service restaurant, customers arrive according to a Poisson process at a rate of one person every minute, and they get to sit at the table after 30 seconds on average.

What is the probability of having 10 customers dining in restaurant at any time?

- **M/M/**∞
- **arrival rate** $\lambda = 1/60$ people/sec
- **service rate** $\mu = 1/30$ people/sec

What is the probability of having 10 customers dining in restaurant at any time?

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \frac{e^{-\lambda/\mu}}{n!}$$

$$P_{10} = \left(\frac{1/60}{1/30}\right)^{10} \frac{e^{-(1/60)/(1/30)}}{10!} \approx 1.63e - 10$$

See slide 9

Problem

Suppose that messages arrive according to a Poisson process at a rate of one message every 3 sec, and that message transmission times are exponentially distributed with mean 2 sec.
The system serves 3 users at a time.

What is the mean time a 4^{th} user will spend in the queue?

- **M/M/m**, with m=3
- **arrival rate** $\lambda = 1/3$ msg/sec
- **service rate** $\mu = 1/2$ msg/sec
- $\rho = \lambda / (m \cdot \mu) = 2/9$

What is the mean time a 4th user will spend in the queue?

$$P_0 = \left[\left(\sum_{n=0}^{m-1}\frac{(m\rho)^n}{n!}\right)+\frac{(m\rho)^m}{m!(1-\rho)}\right]^{-1} = \left[\left(\sum_{n=0}^{2}\frac{(3(2/9))^n}{n!}\right)+\frac{(3(2/9))^3}{3!(1-2/9)}\right]^{-1}$$

$$= \left[\left(\sum_{n=0}^{2}\frac{(2/3)^n}{n!}\right)+\frac{(2/3)^3}{6(7/9)}\right]^{-1} \approx 0.51$$

$$P_Q = \frac{(m\rho)^m}{m!(1-\rho)}P_0 \approx \frac{(3(2/9))^3}{3!(1-2/9)}(0.51) \approx 0.03$$

$$W = \frac{\rho P_Q}{\lambda(1-\rho)} \approx \frac{(2/9)(0.03)}{(1/3)(1-2/9)} \approx 0.03\mathrm{sec}$$

See slides 13 and 15

Problem

Calls arrive to a call-center according to a Poisson process with intensity of 2 calls per minute. The call holding times are exponentially distributed with an average of 5 minutes. Calls that find all operators busy are blocked.

Give the Kendall notation of the system.

How many operators are necessary to keep the blocking probability below 75%?

See slide 19

- M/M/m/m with λ= 2 call/min, μ = 1/5 call/min, $\rho=\lambda/\mu$= 10.

$$B(m,\rho)=\frac{\rho^m / m!}{\sum_{n=0}^{m}\frac{\rho^n}{n!}}\leq 0.75$$

$$B(1,10)\approx 91\% \rightarrow NO$$

$$B(2,10)\approx 82\% \rightarrow NO$$

$$B(3,10)\approx 73\% \rightarrow OK$$

Problem

In a single server - finite buffer system, arrivals can be modeled as a Poisson process with rate 15 s^{-1} and the service times are exponentially distributed with mean 0.05 s.

Give the Kendall’s notation (motivating your choice) and calculate:

a) the average number of customers in the system, assuming the maximum capacity of the system is 5 customers;
b) the blocking probability;
c) the rate of rejected customers.

Kendall's notation: M / M / 1 / N

Poisson arrivals

Exponential service times

One server

Finite buffer

$$\rho = \frac{\lambda}{\mu} = \frac{15}{1/0.05} = 0.75$$

a)
$$E[n] = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} =$$

$$= \frac{0.75}{1-0.75} - \frac{(5+1)0.75^{5+1}}{1-0.75^{5+1}} \approx 1.7$$

See slide 5

b) $$P_N = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}} = \frac{(1-0.75)0.75^5}{1-0.75^{5+1}} \approx 0.07$$

c) $$R_{rejection} = \lambda P_N \approx 15 \cdot 0.07 = 1.05\,\mathrm{s}^{-1}$$

See slides 5 and 6

Problem

In a certain system the arrivals, which can be modeled as a Poisson process, occur at a rate of one every 400 ms, and the service times are exponentially distributed with mean 1 s. Also, there is no limit in the number of servers available.

Give the Kendall's notation (motivating your choice) and calculate:

a) the average number of customers in the system;
b) the average delay per customer;
c) the queuing delay (what do you notice? why is it so?)

Kendall's notation: M / M / ∞

Poisson arrivals

Exponential service times

Infinite servers

$$\lambda = \frac{1}{400 \cdot 10^{-3}} = 2.5\,\mathrm{s}^{-1}, \quad \mu = \frac{1}{1} = 1\,\mathrm{s}^{-1}$$

a) $$E[n] = \frac{\lambda}{\mu} = \frac{2.5}{1} = 2.5$$

b) $$E[\tau] = \frac{1}{\mu} = 1\,\mathrm{s}$$

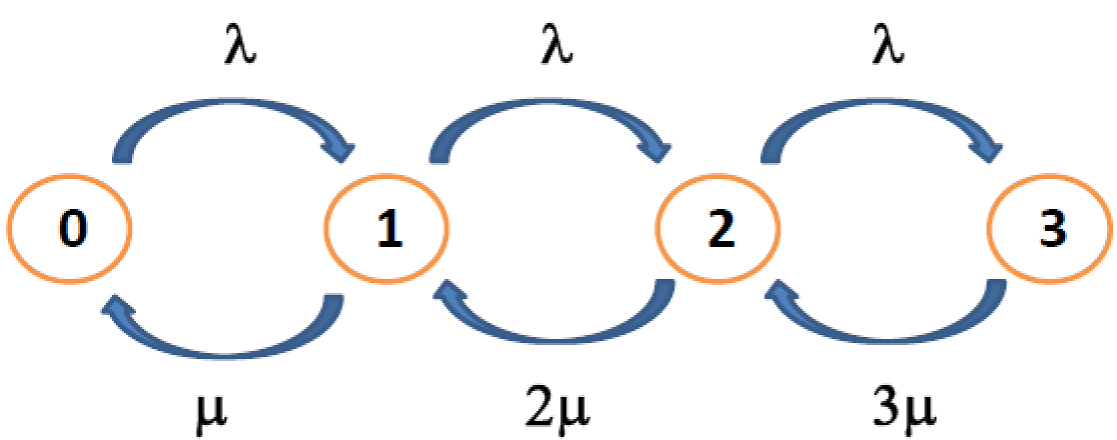
c) As we have an infinite number of servers, there is no queuing delay, i.e.

$$E[\tau_q] = 0\,\mathrm{s}$$

See slide 10

Problem

A queuing system has a Poisson arrival process and exponential service times, and can be represented by the Markov chain depicted below.
Write the Kendall's notation (motivating your choice) and the balance equations for each state.



- M/M/3/3
  - Poisson arriv. → M
  - exp serv. times → M
  - 3 servers
  - no queuing allowed

$$\lambda p_0 = \mu p_1$$

$$(\lambda + \mu) p_1 = \lambda p_0 + 2\mu p_2$$

$$(\lambda + 2\mu) p_2 = \lambda p_1 + 3\mu p_3$$

$$3\mu p_3 = \lambda p_2$$

Problem

Ireland's supporters arrive to a IRFU tickets sales' office according to a Poisson process with intensity of 2 per minute. One employee is serving them with an exponential distribution with an average of 3 per minute. At any time, up to 10 supporters are allowed to be inside the tickets sales' office.

Give the Kendall notation of the system (motivating your choice) and calculate the average number of supporters in the tickets sales' office.

- M/M/1/N with N=10, λ= 2 /min, μ = 3 /min, $\rho=\lambda/\mu$= 2/3
  - Poisson arriv. → M
  - exponentially distr. service → M
  - one employee → 1 server
  - max 10 users allowed in the system

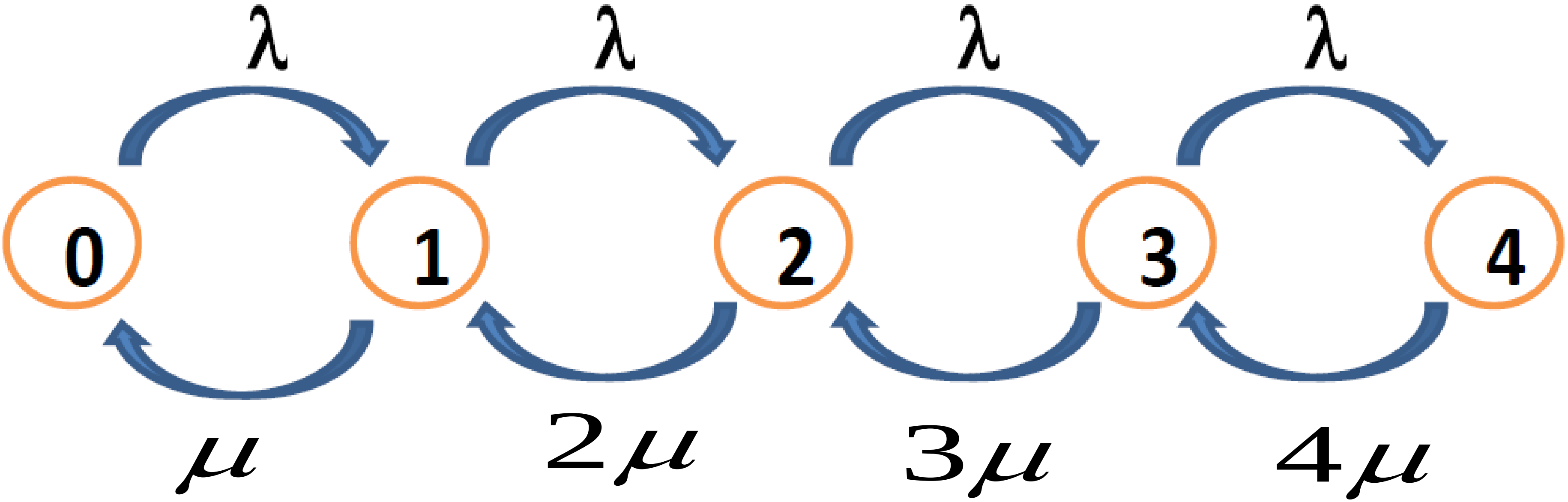
$$E[n] = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} =$$

$$= \frac{2/3}{1-2/3} - \frac{(10+1)(2/3)^{10+1}}{1-(2/3)^{10+1}} = 1.87$$

See slide 5

Problem

A queuing system has exponential service times, and can be represented by the Markov chain depicted



(a) Write the Kendall's notation, giving reasons for your choice and the balance equations for each state assuming that $\lambda = \frac{1}{2}\sec^{-1}$ and the average service time is 3 seconds.

(b) Calculate the blocking probability.

(c) How many servers are needed in order to keep the blocking probability below 0.02?

(a) The Kendall's notation is *M/M/4/4*, since we have a Poisson arrival (*M*), exponential service times (*M*), we can serve up to 4 elements contemporarily (4 servers), and no queuing is allowed (thus we can have at most 4 elements in the system at any time).

The balance equations are:

- for state 0, $\lambda P_0 = \mu P_1$;
- for state 1, $(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2 \Rightarrow \lambda P_1 = 2\mu P_2$;
- for state 2, $(\lambda + 2\mu)P_2 = \lambda P_1 + 3\mu P_3 \Rightarrow \lambda P_2 = 3\mu P_3$;
- for state 3, $(\lambda + 3\mu)P_3 = \lambda P_2 + 4\mu P_4 \Rightarrow \lambda P_3 = 4\mu P_4$;
- for state 4, $4\mu P_4 = \lambda P_3$.

(b) The system utilisation is $\rho = \frac{\lambda}{\mu} = \frac{1/2}{1/3} = 1.5$ and the blocking probability is given by

$$B(m,\rho) = \frac{\rho^m / m!}{\sum_{n=0}^{m} \frac{\rho^n}{n!}} = \frac{1.5^4 / 4!}{\sum_{n=0}^{4} \frac{1.5^n}{n!}} \approx 0.05$$

(c) Therefore with 4 servers, we cannot satisfy the requirement on the blocking probability being $B(m,\rho) < 0.02$. But if we add one more server, we have that

$$B(m,\rho) = \frac{\rho^m / m!}{\sum_{n=0}^{m} \frac{\rho^n}{n!}} = \frac{1.5^5 / 5!}{\sum_{n=0}^{5} \frac{1.5^n}{n!}} \approx 0.01 < 0.02$$

See slide 19